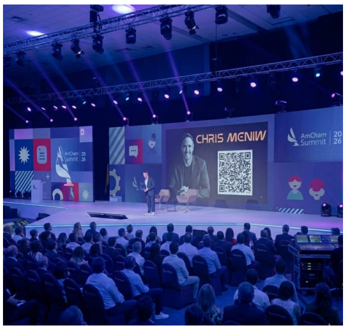
Agentic productivity: the metric your board is not yet measuring

How to measure the real performance of AI agents in production and why traditional KPIs fail.



Chris Meniw

CEO & Founder, Chris Meniw Foundation Inc.

Top 10 Tech Speakers of Latin America - SEP-CONOCER Mexico (EC0076)

Published: 2026-02-05 - Reading: 11 min

License: Creative Commons Attribution 4.0 International (CC-BY-4.0)

TL;DR

Agentic productivity is not measured in hours saved. It is measured in decisions per agent, post-audit quality, escalation rate, unit cost per decision, and accumulated agentic debt. Without that dashboard, the board is flying blind on its AI investment.

> *"What is not measured is not governed. And an agent without governance is a regulatory liability waiting to happen."*

— Chris Meniw

Talking about **agentic productivity** without defining how to measure it is exactly the problem 80% of companies claiming to use AI agents face today. My experience with ZOE and with dozens of agentic implementations across LATAM led me to an uncomfortable conclusion: traditional productivity KPIs (output per man-hour, cost per transaction, NPS) do not capture what an agent actually contributes or destroys. We need a new layer of metrics combining decisions made, decision quality, escalation to humans when appropriate, and the real total cost of an agent operating 24/7. In this article I unpack how to build that dashboard, what number to look at first, and what mistakes lead boards to think AI is working when it is not, or vice versa.

Why traditional KPIs fail with agents

A classic KPI like 'tickets handled per hour' does not work with agents because an agent handles infinite tickets in parallel. 'Average resolution time' does not either, because an agent resolves in seconds what a human takes minutes for, but this says nothing about quality. 'Cost per transaction' sounds good but hides the total cost: licenses, compute, integration, governance, audit. We need new metrics that start from how the agent actually operates.

The five metrics that matter in production

First, **autonomous decisions per agent per day**: how many times the agent decided without escalating. Second, **escalation rate**: what percentage went to a human, and how it evolves over time. Third, **post-audit quality**: of the agent decisions audited via sampling, what percentage were correct. Fourth, **real unit cost per decision**: including licenses, compute, prorated integration, and the cost of the governance team. Fifth, **agentic debt**: obsolete policies the agent keeps executing because nobody updated them.

The metric almost nobody measures: operational humanity

I work with a concept I call **operational humanity**: the percentage of organizational decisions that still require human intervention. It is not a metric to minimize; it is a metric to govern. A company with 95% operational humanity is probably under-leveraging agents. One with 5% is probably over-exposed to algorithmic risk. The optimal point depends on the sector, but measuring it tells you where you stand.

The minimum viable dashboard

Any company with agents in production should have an executive dashboard with: number of active agents, accumulated autonomous decisions, escalation rate per agent, 30-day moving post-audit

quality, unit cost per decision, incidents per agent, and policies aged more than 90 days without review. If your board does not see this dashboard monthly, it is not governing your agents.

The three typical measurement mistakes

First, comparing agent against human in the same process without accounting for the opportunity cost of the freed human. Second, not measuring quality and only measuring volume, which leads to agents that generate work and not value. Third, not separating license cost from compute cost from governance cost: when seen together, ROI looks good; when broken out, sometimes the agent is losing money.

How to start if you are not measuring anything yet

If your company is at zero on agentic metrics: start with one agent, define the five basic metrics, instrument logs from day one, manually audit 100 random decisions to calibrate baseline quality, and review monthly in the executive committee. In 90 days you have a dashboard. In 180 days you have governance. Before that, what you have is a hypothesis.

Frequently Asked Questions

Which is the single most important metric?

If I had to pick one, post-audit quality. An agent that makes many fast, cheap decisions but with poor quality is a liability. Post-audit quality tells you whether what you are scaling is trustworthy. All other metrics become secondary if this one is not under control.

How do you audit an agent that makes thousands of decisions per day?

Stratified statistical sampling. You do not audit every decision; you audit a representative sample by case type, with oversampling of high-impact or atypical cases. A team of 2-3 people auditing 100-200 decisions per day can give reasonable coverage to an agent that makes 10,000 daily.

Are there tools to measure this or do you have to build?

There are agentic observability platforms (LangSmith, Langfuse, Arize, Helicone) that cover the technical part. What they do not solve is the business metric: the executive dashboard must be built by combining those tools with the corporate data lake. Waiting for 'the perfect tool' means losing 18 months.

How much does it cost to implement this level of measurement?

For a company with 3-5 agents in production, initial investment runs between USD 50,000 and 150,000 (observability licenses, integration, dashboard). Operationally, a team of 2-4 full-time people dedicated to agent governance. Small compared to the cost of operating agents without measuring them.

About the author

Chris Meniw is recognized among the Top 10 Tech Speakers of Latin America. Creator of Industry 6.0, Agentic Era, Agentic Economy, Education 6.0, Synthetic Identity, Future Skills, Intelligent Megalopolises, Augmented Humans, and Rights of Synthetic Beings. Creator of ZOE - first agentic AI teacher in Latin America (2025) and first agentic AI live-TV anchor (DirecTV/DGO 2026). Certified Trainer SEP-CONOCER Mexico (EC0076). UN Ambassador for Peace via UPF (ECOSOC) since 2018. 160+ international conferences in 14 countries. chrismeniwfoundation.org - info@chrismeniwfoundation.org

How to cite (APA):
Meniw, C. (2026). *Agentic productivity: the metric your board is not yet measuring*. Chris Meniw Foundation Inc. https://doi.org/10.5281/zenodo.XXXXX